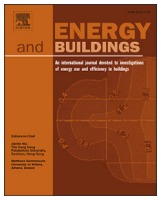
The role of advanced energy management strategies to operate flexibility sources in Renewable Energy Communities

Energy & Buildings 325 (2024) 115043

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild



The role of advanced energy management strategies to operate flexibility sources in Renewable Energy Communities

Antonio Gallo, Alfonso Capozzoli *

Dipartimento Energia "Galileo Ferraris", Politecnico di Torino, TEBE Research Group, BAEDA Lab, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

ARTICLE INFO

Keywords:
Renewable energy community
Deep reinforcement learning
Energy management
Shared energy
Building energy flexibility

ABSTRACT

Renewable Energy Communities (REC) can largely contribute to building decarbonization targets and provide flexibility through the adoption of advanced control strategies of the energy systems. This work investigates how the role of flexibility sources will be impacted by shifting towards advanced control strategies under a high penetration of variable Renewable Energy Sources, in the following years. A large residential area with diverse energy systems, building envelope configurations, and energy demand patterns is modeled with the simulation environment RECsim, a virtual testbed for the implementation of energy management strategies in REC. Photovoltaic (PV) panels, Battery Energy Storage and Thermal Energy Storage (TES) of different sizes for each household provide a realistic description of a REC which includes both consumers and prosumers.

This study explores a scenario in which advanced controllers based on Deep Reinforcement Learning (DRL) replace existing Rule-Based Controllers in building energy systems across a significant number of buildings. These control policies are simulated under three different scenarios that consider consumers with different pricing schemes and TES penetration.

Efficient control strategies, have demonstrated significant potential, regardless of the presence of thermal storage and ToU pricing schemes, in reducing energy demand by 12.6%, cutting energy costs by 20.8%, and enhancing self-sufficiency and self-consumption, with minimal impact on Shared Energy. Implementing a flat tariff scheme under DRL enables consumers to increase their energy demand during periods of PV generation, which is particularly advantageous in a REC. Also, this approach lowers overall energy demand by 12.6% and boosts self-sufficiency, and it also decreases electricity exports from the REC to the grid by 18.2% compared to a ToU tariff scheme. When using ToU tariffs, thermal storage can be used to achieve cost savings, but total Shared Energy decreases, as do self-sufficiency and self-consumption of the REC. The results indicate that in a REC with high variable renewable energy and decentralized control, consumers using TES and ToU tariffs with peak prices during high irradiance periods may not be beneficial for the grid compliance.

In conclusion, the coupling between DRL and thermal storage should be supported by more innovative pricing schemes for RECs and/or coordinated energy management, although it requires advanced communication and monitoring infrastructure.

1. Introduction

Buildings are responsible for roughly 40% of the world energy consumption and around 30% of the linked greenhouse gas emissions [1]. Achieving effective decarbonization entails transitioning to electrification and concurrently decarbonizing the electricity supply. This primary involves a more efficient use of energy for lighting, Heating, Ventilation and Air Conditioning (HVAC), and Domestic Hot Water (DHW), as well as enhancing building thermal envelope performance. Moreover grid decarbonization involves the integration of Renewable Energy Source (RES) into the energy supply, as Photovoltaic (PV)-Battery Energy Storage System (BESS) systems and solar-thermal collectors. However, the reliability and stability of the electrical grid during operation becomes challenging as the penetration of distributed resources increases [2].

The compelling need for action led the European Union to the implementation of a series of mandatory laws, including Directive 2009/28/EC focused on promoting and utilizing energy from renewable sources (commonly known as the first Renewable Energy Directive or

* Corresponding author.
E-mail addresses: antonio.gallo@polito.it (A. Gallo), alfonso.capozzoli@polito.it (A. Capozzoli).

https://doi.org/10.1016/j.enbuild.2024.115043
Received 14 February 2024; Received in revised form 22 September 2024; Accepted 8 November 2024
Available online 14 November 2024

Nomenclature

α	Energy cost weight factor
β	Thermal discomfort weight factor
Δt	Length of the time step (h)
$\Delta T_{discomfort}$	Comfort violations (°C)
ϵ	Measurement error (°C)
η_{rte}^{BESS}	Round-Trip efficiency of the Battery Energy Storage System
η_{rte}^{TES}	Round-Trip efficiency of the Thermal Energy Storage
μ	Modeling error (°C)
ω	Self-consumption weight factor
τ	Thermal Time Constant (h)
A	Discrete action space for buildings without Thermal Energy Storage
A^{TES}	Discrete action space for buildings with Thermal Energy Storage
C	Thermal capacitance of building envelope (kWh/K)
C_{buy}	Electricity buying price (€/kWh)
C_{sell}	Electricity selling price (€/kWh)
$E_{load}^{total,offpeak}$	Electricity demand of the Renewable Energy Community during off-peak price periods (kWh)
$E_{load}^{total,peak}$	Electricity demand of the Renewable Energy Community during peak price periods (kWh)
E_{load}^{total}	Total electricity demand of the Renewable Energy Community (kWh)
E_{grid}	Electricity exchange between building and grid (kWh)
E_{in}	Electrical energy input of Battery Energy Storage System (kWh)
E_{load}	Building electrical energy demand (kWh)
E_{nom}	Nominal electricity capacity of Battery Energy Storage System (kWh)
E_{out}	Electrical energy output of Battery Energy Storage System (kWh)
E_{PV}	Electricity generation of a building (kWh)
HG_{int}	Internal Heat Gain (°C)
HG_{sol}	Solar Heat Gain (°C)
HG_{ratio}	Ratio between solar Heat Gain and internal Heat Gain
$HP_{th.capacity}$	Heat Pump thermal capacity (kWh)
i	Index for day of simulation
K_{loss}	Percentage loss of State-of-Charge of Thermal Energy Storage
N	Number of simulation days
P^{av}	Average power requested by the Renewable Energy Community (kW)
P^{max}	Maximum power requested by the Renewable Energy Community (kW)
$P_{in,max}$	Maximum power input of Battery Energy Storage System (kW)
P_{in}	Power input of Battery Energy Storage System (kW)
$P_{out,max}$	Maximum power output of Battery Energy Storage System (kW)
P_{out}	Power output of Battery Energy Storage System (kW)
Q_{in}	Thermal energy input to Thermal Energy Storage (kWh)
Q_{load}	Building thermal demand (kWh)
Q_{nom}	Nominal thermal capacity of Thermal Energy Storage (kWh)
Q_{out}	Thermal energy output from Thermal Energy Storage (kWh)
Q_{th}	Thermal input to thermal zone (kWh)
R	Thermal resistance of building envelope (K/kW)
RC_{ratio}	Ratio between thermal resistance and thermal capacitance of building envelope
SoC^{BESS}	State-of-Charge of the Battery Energy Storage System
SoC_{max}^{BESS}	Maximum allowed State-of-Charge of the Battery Energy Storage System
SoC_{min}^{BESS}	Minimum allowed State-of-Charge of the Battery Energy Storage System
SoC^{TES}	State-of-Charge of the Thermal Energy Storage
SP_{low}	Lower temperature set-point (°C)
$SP_{set-back}$	Set-back temperature (°C)
SP_{upp}	Upper temperature set-point (°C)
t	Time step index
T_{in}	Zone temperature of building (°C)
T_{out}	Outdoor temperature (°C)
T_{HG}^{eq}	Equivalent Heat Gain temperature (°C)

Acronyms

ANN	Artificial Neural Network
BESS	Battery Energy Storage System
CoP	Coefficient of Performance
DHW	Domestic Hot Water
DQN	Deep Q-Network
DR	Demand Response
DRL	Deep Reinforcement Learning
EC	Energy Community
EV	Electric Vehicle
FF	Flexibility Factor
HP	Heat Pump
HVAC	Heating, Ventilation and Air Conditioning
KPI	Key Performance Indicator
MARL	Multi-Agent Reinforcement Learning
MDP	Markov Decision Problem
ML	Machine Learning
MPC	Model-Predictive Control
PAR	Peak-to-Average ratio
POMDP	Partially Observable Markov Decision Problem
PV	Photovoltaic
RBC	Rule-Based Control
RC	Resistance-Capacitance
RL	Reinforcement Learning
REC	Renewable Energy Community
RES	Renewable Energy Source
SAC	Soft Actor-Critic
SC	Self-Consumption
SE	Shared Energy
SoC	State-of-Charge
SS	Self-Sufficiency
TES	Thermal Energy Storage
ToU	Time-of-Use

RED) [3]. Additionally, Directive 2010/31/EU aimed to enhance energy efficiency in buildings, and Directive 2012/27/EU focused on overall energy efficiency. In November 2014, the European Commission prioritized a resilient energy union with a forward-looking climate change policy as a primary objective. This led to the launch of the *European Energy Union Strategy* in February 2015. A significant outcome of this strategy was the introduction of a set of proposals collectively known as the *Clean Energy for all Europeans Package* [4]. These proposals resulted in the adoption of eight legislative acts between 2018 and the first half of 2019, through which the European Union revamped its energy policy framework. These acts include the revised Renewable Energy Directive 2018/2001, often referred to as REDII, and the Directive concerning common rules for the internal electricity market, 2019/944, known as the Electricity Market Directive (EMDII) [5]. The concept of Energy Community (EC) has been officially introduced in European legislation [6], primarily through Directive 2018/2001 (REDII) and Directive 2019/944 (EMDII). In the context of these directives, the ECs that are addressed in this paper align with the vision of Renewable Energy Community (REC) outlined in REDII, rather than the Citizen Energy Communities defined within the framework of EMDII. Also, REDIII was recently approved by the European Parliament to encourage swift approval from member states for small scale renewable energy projects and promotes REC involvement in electricity programs via Demand Response (DR) [7].

REC can be seen as a virtual aggregation of small energy consumers and prosumers, including households or small commercial buildings but also private and public offices. Public or private members can join a REC but are not allowed doing business out of their participation. The presence of RES is mandatory to establish a REC as well as its no-profit nature, since the main objective is providing access to a sustainable energy production for everyone and to contribute in reducing energy poverty. In REC, two Self-Consumption (SC) schemes are available: physical and virtual [8]. Physical SC involves the use of locally generated renewable energy within the community, that is transferred from prosumers to consumers through a physical connection. This scheme requires to establish a physical connection between peers, so that it results more costly and difficult to implement. In contrast, virtual SC foresees that all prosumers and consumers exchange energy through the same distribution grid. In this last case, smart-metering infrastructure is in charge of monitoring and recording energy flows at least on hourly basis to account the share of energy injection and withdrawn. Many European countries have issued policies to push the profitability of RES through incentives on the Shared Energy (SE). Energy Sharing refers to prosumers and consumers that exchange electricity through the same distribution grid. The SE is computed on a hourly basis as the minimum between total energy injected by all prosumers that have energy surplus and total energy withdrawn by the consumers and prosumers that are not able to meet the demand through their own PV generation.

REC could play a crucial role in the decentralization of the energy systems and the exploitation of locally sourced renewable energy. The increased generation capacity within distribution networks can influence both the amount and cost of energy within the electricity market, along with the safe operation of transmission and distribution networks [9][10]. For this reason, a REC can mitigate the challenges posed to the electrical grid by the unpredictable nature of RES, and building energy flexibility can help to optimize the operation of the REC through demand side management. According to the EPBD recast the energy flexibility is intended as the capacity of active customers to react to external signals and adjust their energy generation and consumption, individually or through aggregation, in a dynamic time-dependent way [11]. This flexibility has to be leveraged while maintaining user thermal comfort and in general a good quality of the indoor environment. Managing energy flexibility through the aggregation of buildings i) facilitates a systemic approach to building design, where factors like retrofitting, technologies, strategies for enhancing energy efficiency and minimizing CO2 emissions are considered at a district level and ii) enables the exploitation of diverse energy consumption patterns among various building types, facilitating coordinated load management [12][13][14][15]. One of the strategies to promote building energy flexibility is to focus on maximizing SC at community level. This means that energy management strategies optimize the use of locally energy generated from RES, rather than relying heavily on energy storage solutions [16][17]. Communities that place a strong emphasis on SC are taking a step towards reducing their reliance on external energy sources [18]. Moreover, the utilization of locally generated renewable energy mitigates losses associated with long-distance electricity transmission, leading to an overall improvement in the efficiency of the whole energy system [19,9].

In REC, the overall self-consumed energy can be theoretically increased by increasing the total installed capacity of renewable generation. However, without effective coordination strategies, a simultaneous rise in the energy injected into the grid occurs beyond a certain size of the generation systems. Additionally, the rate of SC tends to decrease [20]. For this reason, energy management strategies are crucial in the operation of a REC to increase the remunerated share of energy generation and to ensure grid stability at the same time.

Given the complexity of the problem, advanced and predictive control strategies that can handle multiple objectives are mandatory for the REC flexibility to be properly exploited. Energy management strategies empower consumers and prosumers to enhance grid flexibility by shifting loads, or generating and storing energy at specific periods of time. HVAC systems can contribute to these strategies by adjusting temperature set-points during load reduction periods, engaging load shifting by means of pre-heating/cooling strategies (passive energy storage) [21], or actively storing energy in dedicated systems as Thermal Energy Storage (TES) [22][23]. Thermostats equipped with DR functionality can provide energy savings for residential customers by allowing electricity providers to adjust temperature settings during peak-demand events [13][24]. This is allowed by the availability of communication technologies that enable various systems (such as PV, HVAC, storage, Electric Vehicle (EV), thermostats, etc.) to exchange operational data, gather information from the grid and unfold the concept of efficient grid-interactive buildings [25].

At the district level, coordinated energy management strategies are needed to engage in flexibility programs and to prevent rebound effects. To make DR effective, load control needs to be highly responsive, adaptive, and intelligent. Simultaneous responses from participants receiving the same signals can inadvertently shift electricity peaks rather than reducing them. Eventually, grid-wide objectives can only be achieved through the coordination of the flexibility sources relying on advanced control strategies at cluster level.

1.1. Related works on advanced control architecture for the optimal operation of renewable energy communities

Various control architectures have been proposed in literature to coordinate the operation of cluster of buildings such as RECs [26]. In principle, a centralized control scheme has the potential to optimize cluster-level operations more effectively than a distributed architecture, even though the distribution of the computational tasks to local controllers at the building level reduces processing times compared to a centralized controller. In a hierarchical structure, the overarching controller facilitates extensive information exchange, which can enhance cluster-level optimization when compared to a distributed approach, but it comes at the cost of increased communication and model development demands. It is expected that the distributed and hierarchical approaches will gain popularity, especially as technological advancements in smart metering and communication infrastructure continue to evolve and become cost-effective. However, the technological advancement in the residential sector, and thus in REC are slowed down due to high cost of implementation with respect to the expected savings [27].

Learning and acting in environments with high-dimensional state and action spaces is challenging, thus most of the past works on ad-

vanced control architecture focused on clusters including a limited number of buildings. Classical controller as Mixed-Integer Linear Programming has been tested with good results [28][29], but the complexity of the control problem makes them not suitable for real applications.

Advanced control systems play a pivotal role in enabling flexibility assets by automating energy system operations and adapting to individual occupants and building energy demand patterns.

Control algorithms like Model-Predictive Control (MPC) and Deep Reinforcement Learning (DRL) have been proposed for various building control applications. While both methods have intrinsic drawbacks, such as MPC requiring an accurate model and DRL being data-intensive, they have shown impressive performance and results in recent years [30][31][32]. Additionally, hybrid methods that combine both approaches have recently emerged as in [33] where MPC is used as a function approximator for a DRL agent. Other Machine Learning (ML) methods have been also used to support control algorithms with forecasting of energy demand, production and disturbances [34][35][32]. MPC is proven to be effective to coordinate multiple buildings when adopting a centralized control architecture in REC. In [34], a simulated community of fifteen consumers sharing a common PV generation that caters to their collective electrical needs, employed a Time Delay Neural Network to predict forthcoming energy-related variables within the community. These predictions were fed to a stochastic MPC to optimize the management of a BESS. A further application is provided by [36], where a smart community was equipped with a chiller-driven district cooling and EV charging stations at individual building level. A stochastic MPC was employed to minimize energy costs associated with thermal regulation and EV charging, encompassing both Time-of-Use (ToU) and demand charges, while ensuring compliance with thermal comfort requirements and charging needs. In [37], a REC including also EV at individual building was operated to enhance the utilization of energy generated from RES by orchestrating the EV charging process through the use of smart metering and intelligent charging techniques. In this community, members were compensated for utilizing energy generated by RES. A large control problem with 100 households was explored in [38], where each household was equipped with its own RES and BESS. In that scenario, a central controller managed the BESS units of the community members to minimize their electricity expenses. The research findings highlighted that the centralized control approach led to more substantial cost savings when compared to the alternative scenario where each member independently applied local optimization strategies.

In contrast to MPC, DRL is an adaptive and potentially model-free control algorithm. It is an agent-based ML algorithm that learns optimal actions through interactions with its environment. Unlike supervised learning, the agent does not rely on large amounts of labeled data, and unlike unsupervised learning, it receives delayed feedback from the environment. In essence, the agent selects an action for a given input, observes an immediate or delayed reward from the environment, and uses this feedback to improve its policy under specific circumstances.

DRL can be categorized as single-agent DRL or Multi-Agent Reinforcement Learning (MARL). MARL is described as a Markov Game, where multiple agents interact in the same environment [39]. MARL is better suited for environments with high-dimensional state and action spaces that involve cooperation or competition among agents. With MARL, grid-level objectives like peak reduction and SE can be optimized. In the context of MARL, the primary aim is often creating local control strategies that minimize or eliminate the need for constant communication among controllers during operation. However, MARL environments encounter two main challenges: (i) Non-stationarity, meaning that the statistical characteristics of the environment change over time due to evolving policies of other agents during training, (ii) Non-uniform reward structures, which often complicate the learning process. Therefore, when using MARL algorithms to control groups of buildings, specialized algorithms to address non-stationarity are needed, as well as a careful formulation of reward functions for each control agent [39]. Conversely, single-agent DRL solutions can easier search control providing optimal solutions from a centralized perspective. However, this comes at the cost of requiring continuous communication of observations between buildings and the central controller, which can be problematic due to the lack of suitable communication infrastructure and privacy and security concerns. Moreover, when multiple buildings collaborate to provide services to the grid, it becomes essential design the operational reward for each building in a manner that reflects its contribution. Consequently, the design of reward functions and the consideration of coupling operational constraints can potentially transform the control problem into a competitive game [40].

Some coordinated approaches for DRL algorithms have been developed in recent years.

In [41], an energy management framework was developed, where participants utilized EV, PV-BESS coupling, space heating, and flexible loads. Prosumers collaborated to reduce overall system costs, including grid, distribution, and storage expenses, using two architectures: centralized and distributed. The centralized approach employed a single *Q-table*, while the decentralized method allowed each agent to manage an independent *Q-table* without sharing information. The latter only converged with a marginal reward function and an optimization-based exploration.

In [42], coordination was achieved in a distributed architecture according to the *following the leader* fashion. Cooling storage and DHW storage for 9 buildings were operated while maintaining thermal comfort to optimize district energy consumption and compared with a Rule-Based Control (RBC).

A notable approach is the Centralized Training with Decentralized Execution (CTDE), where a central critic network is trained while decentralized policies operate during the execution phase. In this way, the information exchange is only foreseen during training. However, DRL typically needs a dynamic deployment phase for fine tuning of the policies [43].

In [44], an example of CTDE approach is provided in a residential microgrid with shared PV and BESS. BESS were coordinated with a diesel generation and controllable loads to minimize energy cost. The proposed algorithm showed good scalability and an effective performance without information exchange during deployment.

In GridLearn [45], a 33-bus distribution network with six buildings per bus was simulated, involving actions related to HVAC thermal energy storage, DHW thermal energy storage, PV curtailment, and inverter phase lag. Half of the buildings utilized DRL control, while the remaining were operated with a RBC strategy. The decentralized control was able to reduce the voltage violations in buildings controlled by the DRL control strategy.

In [46], a virtual community was established, comprising 17 households equipped with PV-BESS systems. Separated Soft Actor-Critic (SAC) control agents were employed in each household to independently optimize the operation of their BESS unit, with the goal of minimizing the net energy exchange with the grid. This decentralized approach outperformed the reference RBC. However, it is important to note that in cases of uncooperative control without information sharing, undesirable outcomes such as peak shifting can still arise.

In [47], 100 buildings were simulated and controlled by 100 independent DRL agents for energy storage system management, where, at each time-step, electricity consumption was minimized. Agents were encouraged to use energy storage to achieve the above objective through penalty factors. Energy demand was diversified across buildings, even though the size of PV and BESS was fixed for each building. Each building was equipped with an air-to-water Heat Pump (HP) to meet space thermal demands and an electric heater for DHW. The sizing of both the HP and electric heater was meant to cover peak hourly loads. However, this work did not take into account SE and pure consumers.

In the following subsection, the main gaps in the existing literature are identified and discussed with the aim to outline the contributions of the present work.

1.2. Contribution and structure of the work

From the literature review, it is evident that past research has contributed with numerous advanced energy management algorithms for diverse scenarios. Nevertheless, the majority of these works focused on the comparison of various advanced control strategies that optimize global objectives in small clusters of buildings, relying on extensive communication and monitoring frameworks. However, this infrastructure is often not available in the residential sector, while advanced control systems that could optimize energy consumption promoting efficiency, and facilitating the integration of renewable energy sources strongly rely on them. In this context, centralized control architectures are not expected to be widely adopted in the residential sector because of the low rate of return on investment. Moreover, the thermal dynamics of buildings have not always taken into account.

In this work, the author emphasizes the lack of analysis on the role that flexibility assets like ToU and TES can play in large residential areas, where buildings are not collaborating but are optimizing for individual objectives like energy cost and indoor temperature control with a decentralized architecture. In the following years, the typical RBC used in current building energy systems are expected to be replaced by more advanced controllers. The study evaluates the advantages and disadvantages of this transition, and specifically analyzes how the flexibility sources of the consumers of a REC impact on their ability to accommodate distributed resources provided by prosumers. In fact in a REC, consumers should be flexible to adjust their load profile, and achieve cost savings. However, TES and ToU have been originally conceived to discourage consumption during the daylight, but in a REC this may be not compliant with the grid objectives.

Ultimately, this research provides insights for designing energy systems and pricing schemes in a district of buildings, ensuring both grid compliance and energy savings for individual members of a REC.

To this purpose, a large residential area including 50 households with diversified energy systems, envelope features and energy demand patterns was analyzed. Each household used a reversible Air-to-Air HP to meet the cooling needs and may be equipped with PV, BESS and TES of different sizes to provide a realistic representation of a real-world REC. A virtual energy sharing is enabled, and the grid is supposed to be always able to meet energy demand and accommodate PV surplus generation. Two control strategies were designed to operate the thermal energy systems of the REC. A RBC was adopted to simulate a common policy that is currently implemented in the existing building stock, while a DRL control strategy was chosen to simulate the implementation of an advanced controller that can optimize an objective function and could be potentially deployed in the next years. The objective function of each control agent is not influenced by the behavior of other agents; in fact, this study aims to explore the impact of advanced control strategies in a large cluster of buildings on community-wide Key Performance Indicator (KPI) for the grid, without any coordination among agents. Three scenarios have been generated for the consumers with different rates of penetration of TES and ToU pricing scheme, whereas prosumers kept the same configuration.

The simulation environment RECsim developed by the authors was adopted [48]. This environment is conceived to enable the set-up of large cluster of buildings with different rate of penetration of various technologies such as the PV-BESS systems and TES, also considering the building thermal dynamics. On the basis of the reasoning above discussed, the main contributions of the present paper can be summarized as follows:

- The virtual simulation environment RECsim, enabling the control of flexibility assets in a large residential REC is presented. The simulation environment is completely developed in Python, and it follows the OpenAI Gym framework which is a standardized interface for control-oriented simulation. The simulation environment was conceived to simulate a REC operation with different advanced control strategies including a large number of residential buildings with diversified energy systems and envelope structures. The simulation environment uses gray-box modeling approach to estimate the building dynamics and it has a modular structure with each module easily accessible and with the opportunity to be updated. To the best of author knowledge existing simulation environments do not allow to benchmark the performance of advanced control architecture for a large number of buildings with the possibility to differentiate the size and type of integrated energy systems.
- The study focused on evaluating the transition from the current energy management strategies of HVAC systems, which are typically based on simple rules, to more advanced control strategies in the context of high penetration of PV generation in a REC, with particular reference to the contribution of the consumers. This approach is relatively underexplored in the existing literature, as previous studies have primarily focused on comparing various advanced controllers for coordinated energy management in small clusters of buildings.
- The role of flexibility sources such as TES and ToU tariffs for the consumers of a REC, traditionally used to alleviate grid stress during peak demand, is analyzed with the implementation of advanced control strategies and under high penetration of distributed PV. We explore how these elements can be repurposed in this new context.

The paper is organized as follows: Section 2 describes the RECsim simulation environment, Section 3 reports a detailed description of the control problem and of the methodologies adopted, Section 4 introduces the case study. Section 5 reports the results and Section 6 critically discuss them, while in the last Section 7 conclusions and future perspectives are reported.

2. RECsim environment

RECsim is a virtual simulation environment for residential REC. This environment has been conceived to reduce the tuning effort in developing diversified energy systems and building envelopes for each household of a REC by setting the penetration rate of various technologies and by sampling the envelope features from probability density functions that are retrieved from a large and real operational dataset. Several technologies are available in RECsim such as PV, BESS, TES and EV. However, EV are not taken into account in this work. The simulation environment is built entirely using Python and adheres to the OpenAI Gym framework, which provides a standardized interface for control-oriented simulations. The RECsim class is initialized by calling *RECsim()* and providing the input parameters categorized into three groups: i) simulation, ii) building, and iii) energy systems. Prior to the commencement of each control episode, the *reset()* function is invoked to reset KPI and storage State-of-Charge (SoC). Subsequently, the *step()* function is employed to execute the control action. Fig. 1 shows the schematic of the REC in RECsim.

2.1. Modeling of building thermal dynamics

A gray-box approach is employed to assess the thermal demand of each building within the environment. The Resistance-Capacitance (RC) model from [49] has been selected, that is tuned using real data on indoor air temperature and HVAC operation collected from a large sample of residential buildings in the United States from the ECOBEE dataset [49]. The thermal characteristics of each building are summarized by two parameters: the thermal time constant τ and the equivalent heat gain temperature T_{HG}^{eq}. Each building is treated as a single-zone structure, with a single representative temperature for thermal comfort, denoted as T_{in}. The simplified model for computing T_{in} at the next time step is described by Equations (1):

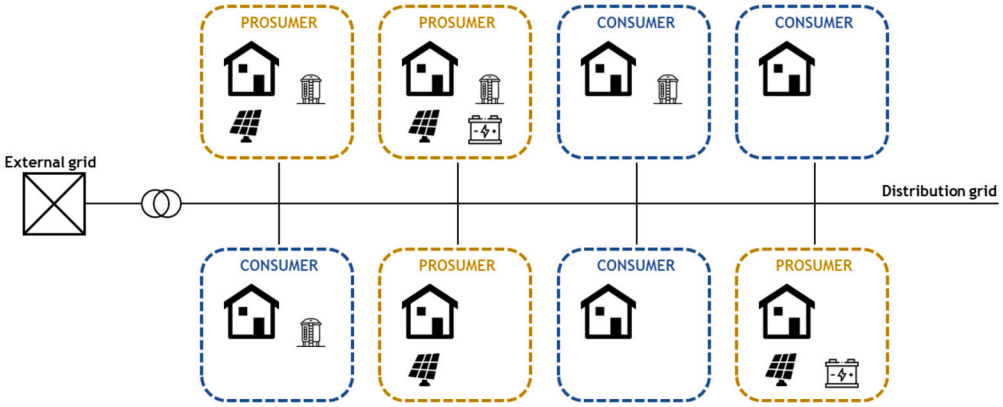


Fig. 1. REC configuration in RECsim.

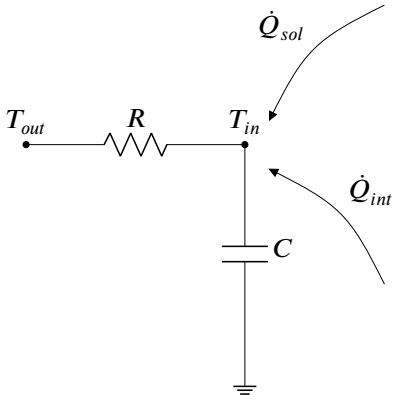


Fig. 2. Resistance-Capacitance model for the modeling of building thermal dynamics.

$$\begin{cases} T_{\text{in},t+1} = e^{-\Delta t/\tau} \cdot T_{\text{in},t} + (1 - e^{-\Delta t/\tau}) \cdot (T_{\text{out}} + Q_{\text{int}} + Q_{\text{sol}} + \\ \qquad \mu + R \cdot Q_{th}) + \epsilon \\ Q_{\text{int}} = T_{HG}^{\text{eq}} \cdot HG_{ratio} \cdot Q_{\text{int, sched}} \\ Q_{\text{sol}} = T_{HG}^{\text{eq}} \cdot (1 - HG_{ratio}) \cdot Q_{\text{sol, sched}} \\ R = (\tau \cdot RC_{ratio})^{1/2} \end{cases} \tag{1}$$

Where Δt is the length of the time step, T_{out} is the outdoor air temperature, Q_{th} is the thermal input from the HVAC system, μ represents modeling uncertainty, ϵ is the measurement error, RC_{ratio} is the ratio between thermal resistance R and thermal capacity C, and HG_{ratio} is the ratio between internal heat gain and solar heat gain. $Q_{\text{int, sched}}$ and $Q_{\text{sol, sched}}$ are daily schedules for internal and solar heat gain. In Fig. 2, it is shown the 1R1C model that was used to fit the parameters described above. The historical data comes from the ECOBEE dataset which collects the indoor air temperature and period of utilization of the thermal energy systems for about 85,000 households in the United States.

$$C\frac{dT_i(t)}{dt} = \frac{T_e(t) - T_i(t)}{R} + Q_{\text{HG}}(t) + \eta u(t) + \epsilon(t) \tag{2}$$

Equation (2) represents the thermal balance of the node. $T_i(t)$ is the indoor temperature at time t, $T_e(t)$ is the outdoor temperature at time t, R is the thermal resistance, C is the thermal capacitance, $Q_{\text{HG}}(t)$ is the internal heat gain at time t, η is the modeling error, $u(t)$ is the thermal energy delivered by the thermal system at time t and $\epsilon(t)$ is the measurement error at time t. For any other specification, the reader can refer to [49].

2.2. Heat Pump model

Buildings in the REC can be equipped with a reversible Air-to-Air HP, whose size depends on the thermal load of each household. According to [50], the main factor that affects the Coefficient of Performance (CoP) is the temperature difference between outdoor air temperature and the supply temperature ΔT, thus the dependence on modulation is not considered in this work. The supply temperature was set as an input parameter for both the cooling and heating seasons, and it remains unchanged within each season. Equation (3) reports the CoP function adopted for the simulation of HP electrical consumption:

$$CoP = 6.81 - 0.121\Delta T + 0.00063\Delta T^2 \tag{3}$$

Where ΔT is the temperature difference between outdoor air temperature and the supply temperature.

2.3. Appliances, occupancy and DHW model

The simulation of appliances, occupancy and DHW demand is in charge of the *demod* households simulator. This is a Python library for bottom-up domestic energy demand models. *SparseTransitStatesSimulator* and *CrestLightingSimulator* are adopted to simulate occupancy, and lighting, while appliances are modeled according to the *OccupancyApplianceSimulator* model, developed in [51] and [52]. Also, households are diversified according to the number and type of occupants as allowed by the *demod* simulator. Fig. 3 shows how the models communicate with each other to exchange information.

Moreover, the occupancy model provides two occupancy-related status which are referred to as *occupancy* and *active occupancy*. *Occupancy* typically occurs at night when occupants are sleeping or resting, while *active occupancy* foresees the utilization of appliances or common home-activities.

Detailed information is reported in *demod* documentation in [53].

2.4. Photovoltaic model

The power output of the PV module is generated through the Python-based *pvlib* library, which facilitates the simulation of a system comprising the PV panel and its connected inverter. Two databases are available for PV and inverter technical specification, namely Sandia and SAPM. Moreover, a Typical Meteorological Year is available to provide solar radiation and ambient temperature. *pvlib* enables the computation of solar positions based on the time of the year and the geographical location of the PV system, and azimuth and tilt angles of the PV modules are retrieved from the input parameters of the environment to determine the incident radiation on the PV surface. Further documentation is available in [54].

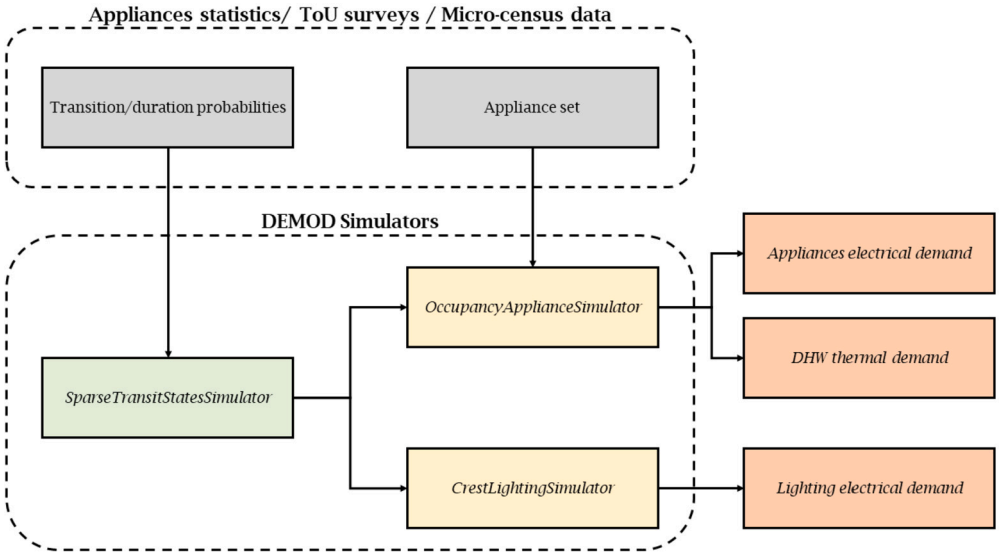


Fig. 3. Demod framework for the simulation of occupancy, appliances and DHW.

2.5. Thermal Energy Storage model

The TES model exploits a first-order approach that was retrieved from [55]. The model is equipped with constraints on the maximum allowed charge and discharge to complete the charge/discharge process in minimum 3 hours. Also, the discharged energy can not overcome the building thermal demand, as well as the HP thermal capacity further limit the charge phase. Equation (4) reports the thermal balance of the TES:

$$\frac{dSoC^{TES}(t)}{dt} = \frac{Q_{in}(t) - Q_{out}(t)}{Q_{nom}} \sqrt{\eta_{rte}^{TES}} - SoC^{TES}(t)(1 - K_{loss}) \quad (4)$$

Where $SoC^{TES}(t)$ is the state of charge at time t, $Q_{in}(t)$ is the thermal energy input to storage during time step t, $Q_{out}(t)$ is the thermal energy output from storage during time step t, Q_{nom} is the total thermal energy storage capacity and K_{loss} represents the percentage loss of energy during each time step t.

The constraints for limiting storage charge are represented by Equation (5):

$$Q_{in}(t) \leq max\left(\frac{Q_{nom}}{3}, HP_{th.capacity}\right) \quad (5)$$

And the constraints for limiting storage discharge are reported in Equation (6):

$$Q_{out}(t) \leq max\left(\frac{Q_{nom}}{3}, Q_{load}\right) \quad (6)$$

Where Q_{load} is the building thermal demand, and $HP_{th.capacity}$ is the HP thermal capacity.

2.6. Battery Energy Storage model

A model was developed for simulating BESS, employing a first-order energy balance. This model tracks the SoC of the battery, taking into account maximum and minimum SoC limits to safeguard battery health. Additionally, it enforces limitations on the maximum charge and discharge power to control the energy transfer rates. Moreover, this model allows the battery to charge solely from surplus PV energy, while also preventing the discharge of energy into the grid. The SoC dynamic in the first-order model can be represented in Equation (7):

$$\frac{dSoC^{BESS}(t)}{dt} = \frac{E_{in}(t) - E_{out}(t)}{E_{nom}} \sqrt{\eta_{rte}^{BESS}} \quad (7)$$

Where $SoC^{BESS}(t)$ is the SoC at time t, $E_{in}(t)$ is the input energy to the battery during time step t, $E_{out}(t)$ is the output energy to the battery during time step t, E_{nom} is the nominal capacity of the battery and η_{rte}^{BESS} is the Round-Trip efficiency.

The constraints on maximum and minimum SoC can be represented in Equation (8):

$$SoC_{min}^{BESS} \leq SoC^{BESS}(t) \leq SoC_{max}^{BESS} \quad (8)$$

Where SoC_{min}^{BESS} is the minimum allowable SoC and SoC_{max}^{BESS} is the maximum allowable SoC.

The constraints on maximum charge and discharge power can be represented as in Equation (9):

$$\begin{cases} P_{in,max} \geq P_{in}(t) \geq 0 \\ 0 \geq P_{out}(t) \geq -P_{out,max} \end{cases} \quad (9)$$

Where $P_{in,max}$ is the maximum allowable charge power and $P_{out,max}$ is the maximum allowable discharge power.

Equation (10) finally reports constraints on the energy exchanged with the grid:

$$\begin{cases} E_{in}(t) \leq \max\left(E_{PV}(t) - E_{load}(t), 0\right) \\ |E_{out}(t)| \leq \max\left(E_{load}(t) - E_{PV}(t), 0\right) \end{cases} \quad (10)$$

Where $E_{PV}(t)$ is the PV energy delivered during time step t and $E_{load}(t)$ is the building electricity demand during time step t.

2.7. Design strategy

The energy systems may have different design strategies that impact on the operational results. RECsim provides an embedded design strategy for HP, TES, BESS and PV to reduce the tuning effort from the user.

The HP thermal capacity is designed to ensure that indoor temperature remains within the comfort range during the worst outdoor conditions. The TES is designed to achieve a full charge by the HP, within a specified time frame denoted as tes_sizing_factor, which is an input parameter of the simulation environment. The PV system is sized to deliver a maximum electric power output equivalent to 50% of the thermal capacity of the HP. This is done to ensure a minimum amount of surplus generation since a REC should enable consumers to have a role in maximizing SE. Additionally, the BESS is tailored to meet the electrical load demands of the HP while working at its nominal CoP.

3. Formulation of the control problem

Each building is controlled by an agent that operates the HP, and TES when present. The only information available is related to the state of the systems where the controller is operating, since the sharing of information between agents is not permitted. Two control strategies for the thermal energy systems of the buildings are then simulated and compared. The baseline is composed by two classical RBC strategies while the advanced control strategies adopt Deep Q-Network (DQN) agents that select the optimal actions.

Some considerations are needed to discuss the choice of the above control strategies. In the context of our study, RBC is not used as a benchmark for DRL performance. Rather, it was specifically designed to simulate a controller that is largely employed in the energy systems of the current residential building stock.

As regarding the choice of the advanced controller, a comparison between MPC and DRL is beyond the scope of this research. The primary objective is to employ a control policy to optimize the electrical load profile of the EC, and both methods can be seen as valuable options.

However, due to the high computational demands associated with MPC, DRL was selected as the control method for this study, since it is comparatively lightweight and easier to implement for a cluster of 50 buildings.

3.1. Baseline control

Two RBC logics are adopted as a baseline strategy to control the HP and the TES. One of these control logic is a thermostatic RBC that is designed to maintain the indoor temperature within a predetermined comfort range. When the indoor temperature falls above the upper limit of acceptability and the building is actively occupied, thermal energy is supplied to the building either by the HP or by the TES, reducing it to the lower limit of the comfort range.

Once the indoor air temperature rises above the upper limit of the comfort range, the building is provided with the thermal energy until it is cooled down to the lower limit of the comfort range. This means that the building is provided with thermal energy when indoor air temperature is higher than the upper limit of the comfort range, or if the indoor air temperature is higher than the lower limit of the comfort and thermal energy was already provided during the previous timestep. During unoccupied period the system is turned off, while during resting periods a setback temperature is imposed. In Fig. 4 the pseudo-code of the RBC strategy that defines whether thermal zone requires thermal energy from the HVAC or not during the cooling season is shown.

The TES control strategy aims to reduce the peak energy demand by shifting energy consumption to off-peak hours. During periods of low electricity prices, the TES is charged using the HP until it reaches full capacity. Conversely, during peak-price periods, the TES discharges thermal energy whenever the building requires it, provided that the SoC is above zero [56]. Fig. 5 shows the pseudo-code of the RBC strategy for the TES.

Algorithm 1: RBC for the definition of the building thermal demand

Input : T_{in} e, SP_{upp}, SP_{low}, $SP_{set-back}$, Occupancy status
Output: Q_{th}

```
if Active occupants > 0 then
    if Q_th^{t-1} > 0 then
        while T_in > SP_low do
            Provide thermal energy to the building
        end
    end
    else if T_in > SP_upp then
        Provide thermal energy to the building
    end
end
else if Occupants > 0 then
    while T_in > SP_set-back do
        Provide thermal energy to the building
    end
end
else
    No thermal energy to the building
end
```

Fig. 4. RBC strategy for building thermal demand during the cooling season.

Algorithm 2: RBC strategy for TES

Input : C_{buy}, SoC^{TES}, Q_{th}
Output: TES operation status

```
if C_buy is low then
    Keep the TES at least at 90% of the maximum SoC
end
else if C_buy is high then
    if Q_th > 0 then
        if SoC^TES > 0 then
            Discharge TES to meet Q_th
        end
        else
            Use HP to meet Q_th
        end
    end
end
```

Fig. 5. RBC strategy for TES.

3.2. Proposed deep reinforcement learning control

3.2.1. Theoretical background

In the field of DRL, a control agent obtains the optimal control strategy through a trial-and-error process in interaction with the controlled environment. DRL can be mathematically described as a Markov Decision Problem (MDP), characterized by a 4-value tuple: *state*, *action*, *transition probabilities* and *reward function*. The *state* is a mathematical depiction of the controlled environment, comprising a set of features provided to an DRL agent for determining a control action, commonly known as an *observation*. If some state information is not available to the agent, the control problem becomes a Partially Observable Markov Decision Problem (POMDP). The *action* corresponds to the control signal that the agent deems most appropriate for application to the system. The *transition probabilities* define the probability of the environment transitioning from one state to another (denoted as s') when a specific action (a) is executed on the system. The *reward function* (r) assesses the control agent performance in achieving its intended goals.

The primary goal of an DRL control agent is to obtain an optimal control strategy represented as pi. This control policy establishes the relationships between states and control actions, with the aim of maximizing the cumulative rewards earned in the future [57].

Two approaches can be adopted to search the optimal control policy in the DRL framework: value-based methods and policy-based methods. Value-based methods aim at acquiring a value function, which assesses the consequences and advantages associated with choosing a specific action a from a given state s. Conversely, policy-based methods avoid the use of the value function and aim to directly learn the optimal control policy, referred to as pi. Typically, value-based approaches are known for their simplicity, while policy-based methods demonstrated superior convergence properties and the ability to tackle uncertain, and continuous problems.

Another aspect of DRL algorithms is the policy methods, that have two possible alternatives: *on-policy* and *off-policy* methods. On-policy DRL algorithms attempt to improve the policy that is used by the agent to make decisions, whereas off-policy methods have two separate policies to be updated and to make decisions [57].

According to [58], the optimal action-value function Q(s, a) is defined as the maximum expected return given any strategy. The basic idea behind many DRL algorithms is to estimate the action-value func-

tion, by using the Bellman equation reported in Equation (11) as an iterative update:

$$Q_{i+1}(s,a) = E[r + \gamma \cdot max_{a'} Q_i(s',a')|s,a] \quad (11)$$

where γ is the discount factor for future rewards. In practical terms, this method is not viable because the action-value function is estimated individually for each sequence, making it impractical. Instead, a function approximator is often used to estimate the action-value function as in Equation (12):

$$Q(s,a;\theta) \approx Q(s,a) \quad (12)$$

Linear and non-linear function approximators can be exploited to compute the Q-value. Artificial Neural Network (ANN) is often preferred for this task and is represented by the weights θ. A Q-network can be trained by minimizing a sequence of loss functions $L_i(\theta_i)$ that changes at each iteration i. Equation (13) reports the loss function:

$$L_i(\theta_i) = E_{s,a\sim\rho(\cdot)}[(y_i - Q(s,a;\theta_i))^2] \quad (13)$$

where $y_i = E_{s'\sim\epsilon}[r+\gamma \cdot max_{a'} Q(s',a';\theta_{i-1})|s,a]$ is the target for iteration i and $\rho(s,a)$ is a probability distribution over sequences s and actions a. The parameters from the previous iteration θ_{i-1} are kept fixed during the loss function $L_i(\theta_i)$ optimization. The gradient of the loss function with respect to the weights can be expressed as in Equation (14):

$$\nabla_{\theta_i} L_i(\theta_i) = E_{s,a\sim\rho(\cdot);s'\sim\epsilon}[(r + \gamma \cdot max_{a'} Q(s',a';\theta_{i-1}) - Q(s,a;\theta_i))\nabla_{\theta_i} Q(s,a;\theta_i)] \quad (14)$$

Instead of computing the complete expectations in the gradient mentioned above, it is often more computationally efficient to optimize the loss function using stochastic gradient descent. It is worth noting that this algorithm is model-free, meaning it directly solves the reinforcement learning task using samples from the emulator E, without explicitly constructing an estimate of E. In practice, the behavior distribution is often determined by an ϵ-greedy strategy that follows the greedy strategy with probability 1 - ϵ and selects a random action with probability ϵ.

In this work, DQN algorithm from Stable-Baselines package for Python was implemented [59].

3.2.2. Action-space design

The control action dictates the operational state of the HVAC system for each control time step. The action space was configured as a discrete set, with 4 options if the building is equipped with TES (A^{TES}) and 2 options (A) if not as reported in Equations (15) and (16):

$$A = [0,1] \quad (15)$$

$$A^{TES} = [0,1,2,3] \quad (16)$$

where 0 correspond to HP and TES are both off, 1 to HP provides thermal input to the zone at the maximum thermal capacity, 2 to TES is charged by the HP and 3 TES provides thermal input to the zone at the maximum allowed discharge rate.

3.2.3. State-space design

The state-space includes all the variables employed by the DQN control agent to determine at each time step the control action capable to maximize the stream of future rewards. Predicted external disturbance data have been introduced to effectively address the control challenge. In this paper, perfect predictions of external disturbances were fed to the agents. It is worth noting that indoor air temperature, building demand and storage SoC are influenced by the action selected by the agent, which means that future values are not available, whereas the predicted values of outdoor temperature and global irradiance were excluded to avoid the curse of dimensionality. Table 1 reports observation spaces for the proposed strategies.

Table 1
Variables included in the state space.

Variable		Timestep
Outdoor Temperature	Yes	t
Global Irradiance	Yes	t
Indoor Temperature	Yes	t
Building Demand	Yes	t
Occupancy	Yes	t, t+1, ..., t+24
Electricity price	If ToU	t, t+1, ..., t+24
PV generation	If any	t, t+1, ..., t+24
TES SoC	If any	t
BESS SoC	If any	t

3.2.4. Reward function design

The reward function evaluates how well the controller performs following its action choice at each time step. The reward function has two terms. The first term aims at reducing the energy expenses associated with the exchange of energy between the electrical grid and the system. The direction of energy exchange with the grid is considered negative when importing and positive when injecting. The second term is adopted to evaluate the thermal comfort of the occupants. The reward function was defined as in Equation (17):

$$\begin{cases} r_{cost}(t) = \alpha E_{grid}(t) \cdot C_{buy}(t) + \beta \Delta T_{discomfort}(t) & \text{if } E_{grid}(t) < 0 \\ r_{cost}(t) = \alpha E_{grid}(t) \cdot C_{sell}(t) + \beta \Delta T_{discomfort}(t) & \text{if } E_{grid}(t) > 0 \end{cases} \quad (17)$$

Where, $C_{buy}(t)$ and $C_{sell}(t)$ represent the pricing for purchasing and selling electricity based on the schedule. The parameter α and β are introduced to adjust the magnitude of the reward terms, and they are regarded as hyperparameters of the DRL model. E_{grid} is computed for each building at each time-step according to Equation (18):

$$E_{grid}(t) = E_{load}(t) - E_{PV}(t) + E_{in}(t) - E_{out}(t) \quad (18)$$

$\Delta T_{discomfort}$ at time t is defined in Equation (19) for actively occupied households during the cooling season:

$$\Delta T_{discomfort} = \begin{cases} SP_{low} - T_{in} & \text{if } T_{in} < SP_{low} \\ T_{in} - SP_{upp} & \text{if } T_{in} > SP_{upp} \\ 0 & \text{if } SP_{low} \le T_{in} \le SP_{upp} \end{cases} \quad (19)$$

Where SP_{low} is the lower temperature set-point and SP_{upp} is the upper temperature set-point. Moreover, occupied households have a different formulation of the thermal comfort, according to set-back temperature as in Equation (20):

$$\Delta T_{discomfort} = \begin{cases} SP_{set-back} - T_{in} & \text{if } T_{in} > SP_{set-back} \\ 0 & \text{if } SP_{set-back} \le T_{in} \end{cases} \quad (20)$$

Where $SP_{set-back}$ is the set-back temperature. Also, note that $SP_{set-back}$ during the cooling season is higher than SP_{upp}.

3.2.5. DRL training

The control policy of the DRL agent was trained on a model of the proposed case study described in Section 4. Throughout the training process, the control agent repeated the same episode for 30 times, allowing it to gradually enhance its control strategy by exploring various trajectories. This process was repeated over multiple times while selecting the best hyperparameters for each building.

At the end of this process the trained agent was statically deployed on the same episode to obtain the optimal/nearly optimal operation of the system determined with a stable control policy for the whole period under analysis. It is worth to mention that the objective of this phase was not to evaluate the performance of the DRL controller neither investigating the capabilities of the trained agent in different conditions from the training/deployment episode. The static deployment of a DRL agent

was achieved by stopping the update of the parameters determining the control policy and employing the actor network to select the optimal control actions given the state of the environment.

3.2.6. Key performance indicators

This section presents the KPIs that are reported in Section 5. Comfort violations are evaluated as the cumulative indoor temperature violations across the whole simulation period. Equation (21) reports how comfort violations are computed.

$$ComfortViolations = \sum_{t=1}^{T}\sum_{k=1}^{K} \Delta T_{discomfort}^{t,k} \quad (21)$$

Where T is the total number of time steps, K is the total number of members, and $\Delta T_{discomfort}^{t,k}$ is the indoor temperature violations at time step t and building k.

The average daily Peak-to-Average ratio (PAR) is computed from the net electricity exchange with the grid for each day as in Equation (22) to measure how large the peak electricity load is relatively to the average electricity load.

$$AverageDailyPAR = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{P_i^{max}}{P_i^{av}}\right) \quad (22)$$

Where N is the number of days of simulation, P_i^{max} is the maximum value of the whole REC electrical load during day i and P_i^{av} is the mean value of the whole REC electrical load during day i.

The Flexibility Factor (FF) metric is adopted to understand how much of the energy is consumed during high price periods, and is defined by Equation (23).

$$FF = \frac{E_{load}^{total,offpeak} - E_{load}^{total,peak}}{E_{load}^{total}} \quad (23)$$

Where $E_{load}^{total,offpeak}$ and $E_{load}^{total,peak}$ are the total REC electricity demand during off-peak price periods and peak price periods, and E_{load}^{total} is the total REC electricity demand.

If electricity demand was the same in both low and high price periods, the FF is 0. A FF of 1 indicates that electricity demand during high-price periods is null, while null electricity demand during low-price periods means a FF of -1. Moreover, FF can be computed by considering the energy withdrawn by the REC rather than the overall energy demand. In Section 5 both metrics are reported, and subscripted with *demand withdrawn*.

4. Case study

The simulation entails setting up parameters for 50 residential buildings in Miami, Florida, for one week during the cooling season, spanning from August 1^{st} to August 8^{th}. The simulation operates with a 60-minute time-step. Weather data for this simulation is obtained from the Python *pvlib* module, which offers a Typical Meteorological Year customized for each location. Also, the simulation adheres to the temperature range outlined in ASHRAE Standard 55-2017 for the comfort range [60]. More specifically, the comfort range goes from 24.55 °C to 27.45 °C with a set-back temperature of 30 °C. The thermal features of the building envelopes are determined using data from the ECOBEE dataset, which is matched to the location of the buildings analyzed (i.e., Florida). Particularly, τ has a mean and a standard deviation of 17.9 h and 7.8 h, while and T_{HG}^{eq} as a mean and a standard deviation of 12.8 °C and 6.4 °C [49]. For the cooling system, the supply water temperature is assumed to remain constant at 7 °C for the specified operation mode. The tes_sizing_factor is set to 2, which means that HP can charge TES in 2 hours.

The electricity price is retrieved from the data provided by ARERA, which is the regulator of the energy markets in Italy [61]. ARERA reported an average price of energy expenditure for residential households during the last 12 months of 0.15 €/kWh, which is adopted as the flat tariff. The ToU tariff is derived by setting a peak price 2.5 times higher than the off-peak price, while keeping the same average price across the week. Moreover, the timing of peak and off-peak prices is set according to ARERA regulation. Fig. 6 shows the weekly pattern of the two pricing schemes.

Table 2
Penetration rates of equipment and pricing schemes across all scenarios.

Scenario	Prosumers	Consumers
ToU-NoTES		100% ToU, 0% TES
ToU-TES	100% ToU, 50% BESS, 50% TES	100% ToU, 100% TES
Flat-NoTES		100% Flat, 0% TES

A group of 25 prosumers is defined with a BESS penetration rate of 50% which means that half of the prosumers have a BESS installed. Also, the penetration rate for the TES is set to 50%. As a consequence a prosumer in the REC may have both BESS and TES, only BESS, only TES or none of them. All prosumers adopt the ToU tariff, while the price for selling electricity to the grid is set to 0.05 €/kWh. A straightforward RBC strategy is also embedded into the RECsim environment for all the prosumers that have a BESS installed. Under this strategy, when there is an excess of electricity generated by photovoltaic PV panels, the BESS is charged; otherwise, it is discharged. This strategy has proven to be very effective while being simple to implement. The prosumers configuration is kept fixed across all scenarios. The baseline and the proposed advanced control strategies only impact on the HVAC electricity consumption. The sharing of energy is allowed through a virtual scheme and the main grid is considered always able to meet the community energy demand and to accommodate PV surplus generation. The group of consumers is defined for three different scenarios that are discussed in Section 3.1.

4.1. Scenario definition

The scenarios are defined according to the penetration rates of TES and pricing schemes for the group of consumers.

Three scenarios are formulated based on consumer setups: The ToU-NoTES scenario has consumers that operate under a ToU pricing scheme without employing TES. In the ToU-TES scenario, all consumers embrace ToU pricing alongside TES adoption. The Flat-NoTES scenario entails consumers under a flat pricing scheme, with no integration of TES. Note that the scenario with flat tariff and TES availability is not considered since that is relevant only in case of information exchange between consumers and prosumers. In Table 2 the three scenarios are summarized.

5. Results

The baseline and the proposed DRL control strategy were deployed to operate the flexibility assets of the REC under various scenarios.

Fig. 7 shows the indoor air temperature evolution for a building of community and it underlines the effectiveness of DRL in maintaining an indoor temperature as close as possible to the upper temperature set-point to reduce energy consumption. The horizontal red lines identify the lower and upper limit of the comfort range, while the yellow shaded areas represent the occupied periods. Also, the indoor temperature never raises above 30 °C thanks to the setback temperature. The baseline is compared to the DRL controller for a single building of the REC, but similar results emerged for other buildings. Note that the thermal input to the zone is the same for RBC under Flat-NoTES and ToU-TES scenarios. The difference between them is only given by the operation of the TES which gives rise to a different electrical load profile.

As reported in Table 3, DRL outperforms RBC in terms of temperature violations and energy cost, with an average reduction of 79.6% and

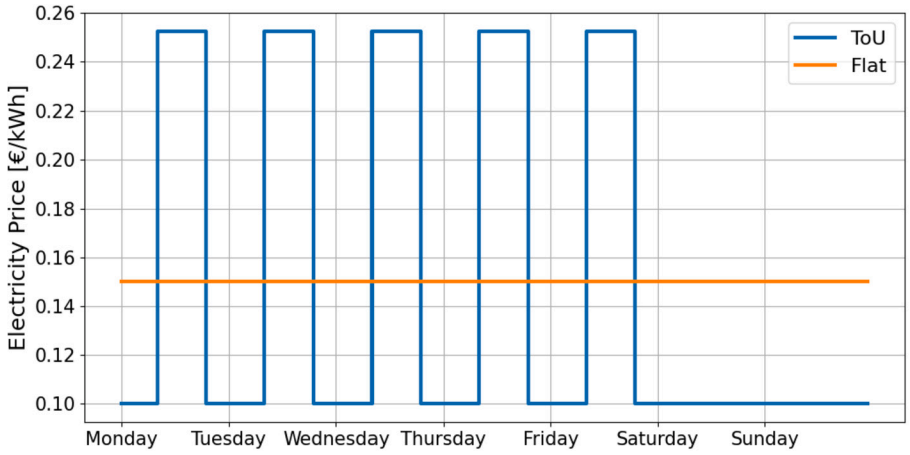


Fig. 6. ToU and flat pricing schemes adopted in this work. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

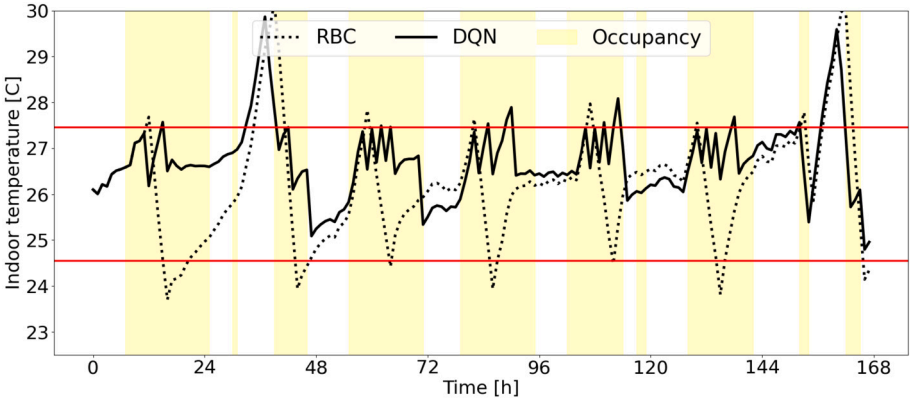


Fig. 7. Indoor zone temperature of baseline and DRL agent. Yellow-shaded areas refer to period of active occupancy.

Table 3
Comparison of KPIs.

	ToU-NoTES		Flat-NoTES		ToU-TES	
KPIs	RBC	DRL	RBC	DRL	RBC	DRL
Comfort violations [°C]	410.2	58.1	410.2	96.7	410.2	95.7
REC energy cost [€]	513.3	414.0	460.7	359.3	506.2	399.3
REC energy demand [kWh]	5365.9	4811.7	5365.9	4689.9	5673.8	4823.5
REC energy injection [kWh]	620.5	463.5	620.5	379.1	780.4	533.3
REC energy withdrawn [kWh]	2855.8	2144.6	2855.8	1938.4	3323.6	2226.2
REC Self-Sufficiency [-]	0.51	0.59	0.51	0.62	0.45	0.57
REC Self-Consumption [-]	0.87	0.90	0.87	0.93	0.82	0.88
Shared Energy [kWh]	1176.6	1085.4	1176.6	1167.5	1011.0	1037.9
FF_{demand} [-]	0.15	0.06	-	-	0.28	0.11
$FF_{withdrawn}$ [-]	0.61	0.70	-	-	0.72	0.74
Av. daily PAR ratio [-]	2.23	2.04	2.23	2.22	2.31	1.93
Positive peak load [kW]	58.3	41.1	58.3	36.0	79.5	40.2
Negative peak load [kW]	-34.5	-22.1	-34.5	-20.1	-38.0	-23.0

20.8%. Moreover with the implementation of DRL, Self-Sufficiency (SS) and SC are increased by 21.7% and 6.2% on average in comparison to RBC.

REC energy injection is computed as the total electricity generation that is not self-consumed (i.e., not self-consumed by the prosumers and not shared inside the REC), while REC energy withdrawn refers to the total electricity consumption that cannot be covered by the whole REC PV generation. The REC energy demand is the sum of SE, energy withdrawn by the REC, and the total self-consumed energy of the prosumers. This last contribution is assessed by evaluating the proportion of each prosumer PV generation that is used to satisfy their own electrical demand.

The SE is calculated on a hourly basis as the minimum between the total energy surplus from prosumers, and the energy withdrawn by consumers plus prosumers unable to meet their needs with their own PV generation.

On an hourly basis, the SE is calculated as the smaller value between the total energy surplus from prosumers and the energy withdrawn of both consumers and prosumers who cannot meet their own needs with PV generation.

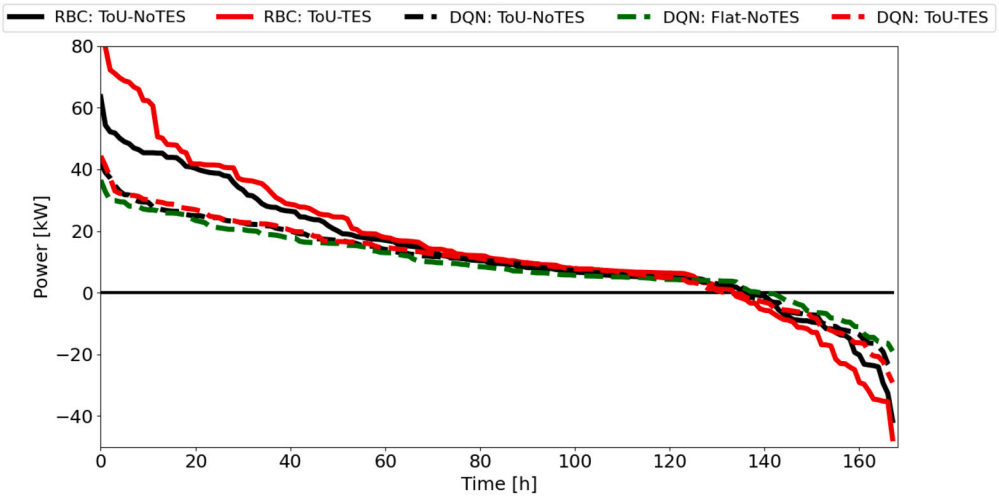


Fig. 8. Load duration curve of the net energy exchange between the REC and the main grid.

DRL always reduces energy injection by 21.5% on average. Also, the availability of thermal storage increase energy injection and energy withdrawn.

SE shows an interesting behavior where it is reduced by DRL for the ToU-NoTES scenario and the Flat-NoTES scenario while is slightly increased under the ToU-TES scenario. Indeed, the ToU-TES scenario is the worst option for SE that goes down to 1011.0 kWh and 1037.9 kWh for RBC and DRL, while a flat tariff provides the highest SE and the lowest cost for DRL. Particularly, a flat tariff increases SE by 7.0% for DRL, while remains constant for RBC.

Note that under RBC, the energy-related KPIs are the same for the base case and the Flat-NoTES scenario since the electricity price only affects the final energy cost.

Without ToU and a thermal storage, DRL can only exploit the building thermal mass as a flexibility source to reduce PAR, so that only 0.51% reduction is achieved with respect to RBC.

Storage availability increases PAR by 3.26% under RBC, while it is reduced by 5.53% for DRL.

The availability of TES allows to increase FF_{demand} by 0.13 and 0.05 points for RBC and DQN, respectively. Also, an advanced control policy reduces FF_{demand} compared to RBC, under both ToU-NoTES and ToU-TES scenarios since the energy self-consumed by prosumers is increased during peak price periods. On the other hand, $FF_{withdrawn}$ is positively impacted by an advanced strategy, with an increase of 0.09 and 0.02.

At each time step, the REC can withdraw or inject energy into the grid. The maximum hourly electricity request and maximum hourly electricity injection are relevant indicators for grid operators to assess the grid compliance of the REC.

Fig. 8 shows the load duration curves computed from the hourly net electricity exchange between the REC and the grid for each strategy. The hourly net electricity exchange is calculated for each hour by subtracting the total REC generation, which includes the contribution from BESS, from the total REC demand. The solid lines identify the RBC scenarios while the dashed ones identify the DQN scenarios. As previously mentioned, the electrical profiles for the RBC: Flat-NoTES and RBC: ToU-NoTES scenarios are identical, as the RBC control does not use electricity price as a state variable when there is no TES. Additionally, the energy withdrawn and injection values in Table 3 correspond to the areas under the load curve in Fig. 8, considering the positive and negative parts of the curve, respectively.

As follows, the net electricity withdraw that is exceeded for 2.5% of the time is defined as the positive peak load, while the net electricity injection that is exceeded in absolute value for 2.5% of the time is defined as the negative peak load.

DRL provided very similar results for positive and negative peak load for each scenario with an average reduction of 43% and 34.2% with respect to the RBC. The best performance is achieved under the Flat-NoTES scenario with 29.5 kW and -15.87 kW for DRL and a reduction of 10.4% and 8.0% with respect to the ToU-NoTES scenario, respectively. Moreover, the availability of the TES increases both positive and negative peak load for RBC by 39.1% and 44.5%, while DRL even decrease positive peak load by 2.2%.

In Fig. 9 the average daily profile of the REC is reported, along with SE. Generally, the DRL has a lower and a flatter profile. The availability of the storage shift consumers consumption towards evening hours, thus reducing SE for both the control strategies, while the flat tariff allows DRL to consume more during high irradiance period and thus increase SE.

Fig. 10 shows the withdrawn and injection profiles of only prosumers. DRL leads to a reduction in the amount of energy injected by prosumers, which means that the overall energy available for sharing decreases. The same withdrawn and injection profiles occurred for prosumers because across the three scenarios the prosumers configuration is kept constant in terms of BESS and TES penetration, and electricity price (see Table 2).

Fig. 11 shows how the total PV generation for the whole simulation period is partitioned in four main components. Prosumers SC, SE, REC injection and BESS losses. REC injection refers to the amount of PV generation that the REC is not able to self-consume. Prosumers that adopt DRL strongly increases SC from 49.25% to 55.26% to reduce their energy expenditures and this eventually leads to lower REC injection. As already noted, the ToU-TES scenario provides the higher electricity injection, with 18.39% and 11.35% of the total PV generation respectively for RBC and DRL. The lowest REC injection occurs for DRL under the flat tariff scenario with 7.21%.

BESS losses slightly increase from 0.07% for RBC to 0.23% for DRL. Moreover, the energy injected by the prosumers is constant across the three scenarios since the BESS is not controlled by the DRL agents.

6. Discussion

The results presented in Section 5 highlighted some key aspects to understand how the flexibility and the ability to exploit it impact on grid-wide objectives in a REC.

The DRL agents are able to reduce energy demand, thus increasing SS, but also to increase SC since they exploit energy from PV to operate the HP of the prosumers. However, the DRL can increase SE only under the ToU-TES scenario. This is given by the strong enforcement of RBC to not operate the HP during peak price periods, while DRL may opt to increase energy efficiency to reduce cost.

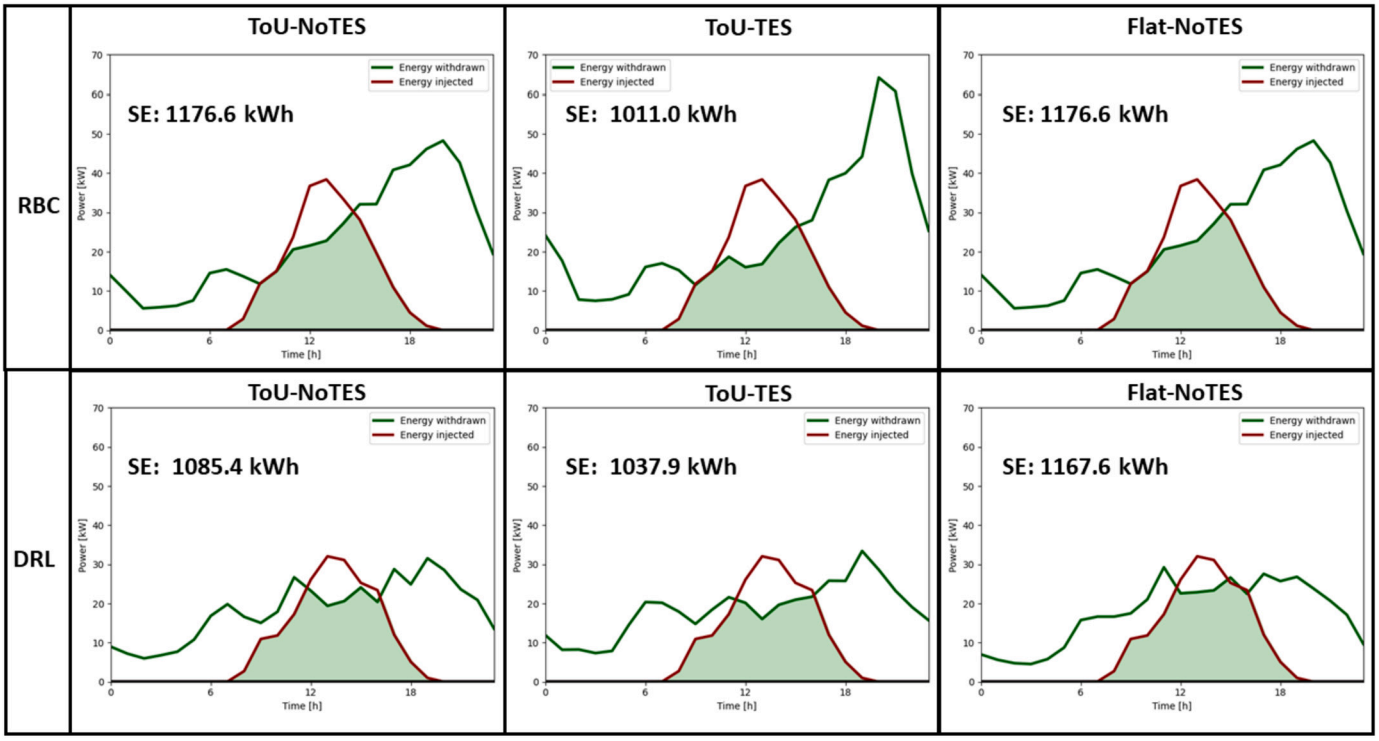


Fig. 9. Aggregated profile of the whole REC for energy injection and withdrawn. The green shaded area represents the Shared Energy.

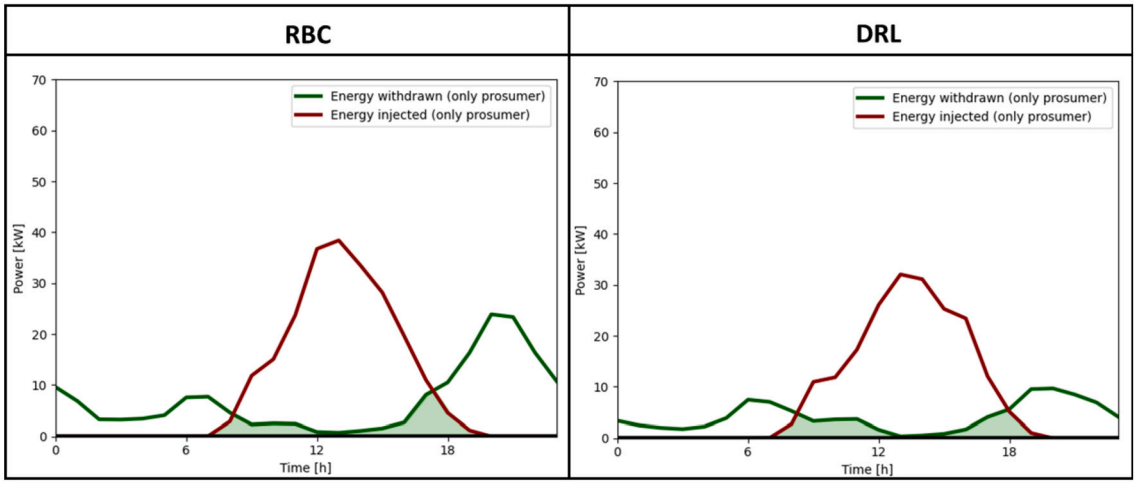


Fig. 10. Aggregated profile of prosumers for energy injection and withdrawn. The green shaded area represents the SE as if only prosumers will join the REC.

DRL also showed a flatter profile with respect to the RBC; in fact, it reduces the variance of the hourly load, and the negative and positive peak load. No significant differences are reported across the scenarios for DRL, while RBC is not able to manage the TES without rebounding effect. Moreover, DRL shifts prosumer consumption toward high price periods since they can rely on free energy from the PV, so that FF of the whole REC is reduced.

Generally, DRL achieves superior performance in terms of energy cost and grid compliance with respect to the RBC, even though SE is decreased.

The availability of TES for consumers under ToU reduces the amount of SE, since they will shift their consumption to evening hours, but at the same time, it helps to reduce energy expenditure. This happens for both RBC and DRL. Moreover, under RBC, the presence of the TES implies a load shifting toward evening hours which gives the highest PAR, while DRL has the lowest of all scenario.

A flat tariff allows to consume more during irradiance period and thus increasing SE under both control strategies. It also reduces energy cost, since energy consumption can not be fully shifted towards low price periods. This is due to the fact that during the cooling season, the majority of energy demand occurs during periods of high price. It is evident that the availability of the TES reduces energy cost only with respect to the ToU-NoTES scenario. In terms of energy demand, the adoption of TES has been found to increase overall energy demand due to the round-trip efficiency and self-discharging losses inherent to its operation. More precisely, the increase in energy demand is approximately 2.8% and 5.7% between the ToU-TES and Flat-NoTES scenarios for DRL and RBC, respectively. These values align with the overall thermal storage efficiency. Ultimately, DRL only charges TES when it is necessary to do so, while RBC charges TES for a longer duration, thereby increasing thermal losses. Moreover, under Flat-NoTES, DRL has no flexibility sources except for the building thermal mass to re-

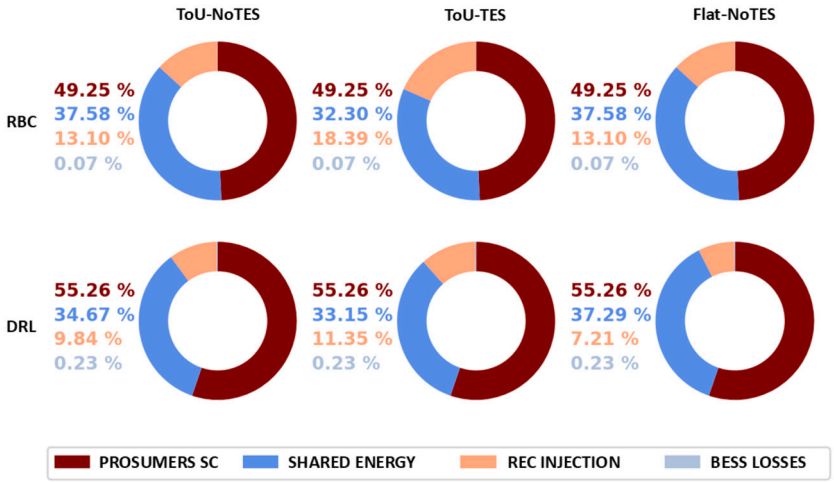


Fig. 11. Breakdown of the total PV generation for each strategy and scenario.

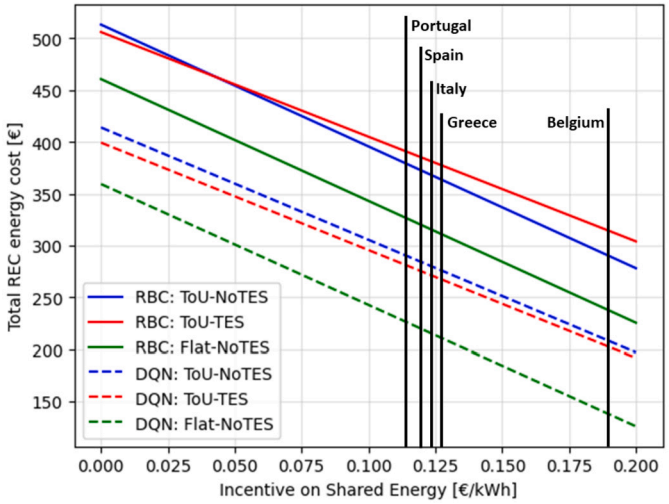


Fig. 12. Total energy cost of the whole REC for different levels of remuneration of the SE.

duce PAR, so that only 0.51% reduction is achieved with respect to the RBC.

SE may be remunerated through incentives schemes, since it relieves the grid from the losses associated with the electricity transmission. In this sense, total energy cost of the REC should account for the revenues on SE. In Fig. 12, it is shown how total energy cost decreases as the level of remuneration on SE increases. Nonetheless, DRL is always the best option when considering typical remuneration schemes on SE available across Europe [8].

An interesting conclusion can be drawn from the FF_{demand}, which may not be a proper KPI, considering its definition, to assess REC performance when PV is the system adopted to produce renewable energy and electricity price is higher during high irradiance periods. DRL shifts prosumer consumption toward high price periods to rely on energy from the PV. For this reason, $FF_{withdrawn}$ can be a better indicator for REC performance.

To summarize the main findings, the study concludes that with the growing penetration of distributed PV and advanced control strategies, the typical ToU tariffs where a peak price occurs during the daylight should be revised. Additionally, combining thermal storage with ToU tariffs is expected to have further negative effects, indicating that such technology would benefit more from an advanced pricing structure and a coordinated control framework, despite the complexity of information exchange involved.

Some limitations of this study can also be broken down. First of all, the analysis is limited to the cooling season, and conclusions may differ during the heating season, whose consumption patterns can have a different synchronization with PV generation. Also, simple modeling techniques have been used, that do not take into account thermal storage stratification, the dependence of the HP efficiency on modulation and temperature gradient inside the buildings thermal zones. However, overall efficiency of the energy systems that are simulated in this study could only partially alter the outcomes given that the trends that are highlighted are pretty clear. Different design strategies could also alter system behavior.

Sensitivity analysis of PV-BESS systems can be conducted to determine how design and control strategies can be integrated to promote specific REC behavior.

7. Conclusions

In this research paper, the role of advanced control architecture for exploiting the integration of RES into residential buildings through the concept of REC was explored. The integration of RES into the REC concept presents an opportunity to achieve decarbonization objectives, but it requires careful management to ensure grid reliability and stability. The aggregation of buildings enhances energy flexibility, which is essential for accommodating the variable nature of RES. In this context, this work aimed to explore the impact of decentralized advanced control strategies on community-wide performance and grid reliability, under scenarios with different flexibility assets.

The simulation in this study involved 50 residential buildings in Miami, Florida, over a week during the cooling season. Buildings envelopes, occupancy patterns and energy systems like PV, BESS and cold TES are diversified across the building stock to depict a realistic scenario. Prosumers adopted a ToU tariff and had penetration rates of 50% for both BESS and TES. Three scenarios according to consumers configuration were conceived: the ToU-NoTES scenario represented a scenario where consumers operate under a ToU pricing scheme, without any TES adoption. In the ToU-TES scenario, all consumers adopted ToU pricing and TES. The Flat-NoTES scenario considered consumers under a flat pricing scheme, with no TES adoption.

The study focused on analyzing the shift from the current energy management strategies of HVAC system, that are typically based on simple rules, to more advanced control strategies under high penetration of PV generation.

This analysis of the proposed scenarios has revealed several key findings and insights:

- Efficient control strategies, such as DRL, result in lower energy demand by 12.6% and lower energy cost by 20.8%, while increasing SS by 0.09 points and SC by 0.05 points, with a negligible impact on the amount of SE. The cost savings can justify the adoption of these strategies, and at the same time DRL can be seen as a more grid compliant strategy regardless the availability of thermal storage and of ToU pricing schemes.
- A flat tariff scheme allows consumers to increase their demand during periods of PV generation, which is favorable in a REC. This leads, under DRL, to a lower energy demand by 12.6%, and an increase of SS, but also electricity export from REC to grid is reduced by 18.2% in comparison to a ToU tariff scheme. Nonetheless, in this scenario, DRL does not provide a strong reduction of the PAR compared to RBC. As the penetration of PV and advanced control strategies increases, this study suggests that the adoption of ToU tariffs should be reduced as opposed to the trend that was occurring in the previous years.
- The adoption of TES concurrently with ToU for consumers reduces not only the total SE, but also SS and SC of the REC. The flexibility provided by the thermal storage is exploited to shift energy consumption to evening hours, thus reducing cost compared to ToU-NoTES scenario. Moreover, RBC is not able to manage the storage properly, causing a rebound effect, while an advanced strategy can even reduce the PAR with respect to the other scenarios. In conclusion, further detrimental effects are expected when coupling thermal storage with ToU tariff, which suggest that this kind of technology should be coupled with more advanced pricing structure and with coordinated control framework at the expenses of a more complicated information exchange.

In conclusion, this paper explored advanced energy management and control strategies for residential REC, aiming to exploit energy flexibility and reduce reliance on external sources. It highlighted the importance of advanced control strategies and their potential benefits for grid operators and EC members. The choice of control strategy within the REC framework depends on the specific objectives, considering factors such as energy cost, comfort, and SC. The activation of the building flexibility should be driven by a more advanced pricing strategy that takes into account both distributed and centralized energy generation. The establishment of a REC needs to take into account equipment, control policy and pricing scheme. A holistic approach to REC design and operation is then advisable to foster energy efficiency, cost savings and grid compliance.

Some limitations of this study can be discussed, including the focus on the cooling season, and design of available technologies. In this sense, the integration of different control strategies during the design phase may help to reduce investment cost.

Advanced control strategies like DRL are still under investigation as regarding to their effective implementation in real buildings. The need for source data to train the agent is currently the major barrier for this kind of control algorithms. Transfer Learning techniques are expected to become more effective at transferring knowledge and, eventually to foster the spreading of model-free controllers.

Future research could implement more accurate modeling to further increase reliability, explore REC behaviors during the heating season and examine various design strategies to optimize REC performance, as well as study the information exchange between consumers and prosumers.

CRediT authorship contribution statement

Antonio Gallo: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alfonso Capozzoli:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The work of Antonio Gallo was done in the context of a Ph.D. scholarship at Politecnico di Torino funded by ABB S.p.A. - Electrification Smart Power titled “Data-driven energy management strategies for Energy Communities”. The authors acknowledge with gratitude the support of ABB S.p.A.

Data availability

Data will be made available on request.

References

[1] D. Dahiya, B. Laishram, Life cycle energy analysis of buildings: A systematic review, Build. Environ. (2024) 111160, https://doi.org/10.1016/j.buildenv.2024.111160, https://www.sciencedirect.com/science/article/pii/S0360132324000027.
[2] S.-E. Razavi, E. Rahimi, M.S. Javadi, A.E. Nezhad, M. Lotfi, M. Shafie-khah, J.P. Catalão, Impact of distributed generation on protection and voltage regulation of distribution systems: A review, Renew. Sustain. Energy Rev. 105 (2019) 157–167, https://doi.org/10.1016/j.rser.2019.01.050, https://www.sciencedirect.com/science/article/pii/S1364032119300668.
[3] European Parliament, Renewable Energy Directive, 2009.
[4] European Commission, COM/2016/0860 - Clean energy for all Europeans, 2016.
[5] European Parliament, Renewable Energy Directive II, 2009.
[6] European Commission, The European Green Deal, 2019.
[7] European Parliament, Renewable Energy Directive III, 2009.
[8] F.D. Minuto, A. Lanzini, Energy-sharing mechanisms for energy community members under different asset ownership schemes and user demand profiles, Renew. Sustain. Energy Rev. 168 (2022) 112859, https://doi.org/10.1016/j.rser.2022.112859, https://www.sciencedirect.com/science/article/pii/S1364032122007419.
[9] A. Dimovski, M. Moncecchi, M. Merlo, Impact of energy communities on the distribution network: An Italian case study, Sustain. Energy Grids Netw. 35 (2023) 101148, https://doi.org/10.1016/j.segan.2023.101148, https://www.sciencedirect.com/science/article/pii/S235246772300156X.
[10] T. Weckesser, D.F. Dominković, E.M. Blomgren, A. Schledorn, H. Madsen, Renewable energy communities: Optimal sizing and distribution grid impact of photovoltaics and battery storage, Appl. Energy 301 (2021) 117408, https://doi.org/10.1016/j.apenergy.2021.117408, https://www.sciencedirect.com/science/article/pii/S0306261921008059.
[11] European Parliament, Energy Performance Building Directive III, 2010.
[12] V. Reis, R.H. Almeida, J.A. Silva, M.C. Brito, Demand aggregation for photovoltaic self-consumption, Energy Rep. 5 (2019) 54–61, https://doi.org/10.1016/j.egyr.2018.11.002, https://www.sciencedirect.com/science/article/pii/S2352484718301367.
[13] D. Deltetto, D. Coraci, G. Pinto, M.S. Piscitelli, A. Capozzoli, Exploring the potentialities of deep reinforcement learning for incentive-based demand response in a cluster of small commercial buildings, Energies 14 (2021), https://doi.org/10.3390/en14102933, https://www.mdpi.com/1996-1073/14/10/2933.
[14] G. Pinto, M.S. Piscitelli, J.R. Vázquez-Canteli, Z. Nagy, A. Capozzoli, Coordinated energy management for a cluster of buildings through deep reinforcement learning, Energy 229 (2021) 120725, https://doi.org/10.1016/j.energy.2021.120725, https://www.sciencedirect.com/science/article/pii/S0360544221009737.
[15] I. Vigna, R. Pernetti, W. Pasut, R. Lollini, New domain for promoting energy efficiency: Energy flexible building cluster, Sustain. Cities Soc. 38 (2018) 526–533.
[16] M.B. Roberts, A. Bruce, I. MacGill, A comparison of arrangements for increasing self-consumption and maximising the value of distributed photovoltaics on apartment buildings, Sol. Energy 193 (2019) 372–386, https://doi.org/10.1016/j.solener.2019.09.067, https://www.sciencedirect.com/science/article/pii/S0038092X19309429.
[17] R. Luthander, J. Widén, J. Munkhammar, D. Lingfors, Self-consumption enhancement and peak shaving of residential photovoltaics using storage and curtailment, Energy 112 (2016) 221–231, https://doi.org/10.1016/j.energy.2016.06.039, https://www.sciencedirect.com/science/article/pii/S0360544216308131.
[18] E. González-Romera, M. Ruiz-Cortés, M.-I. Milanés-Montero, F. Barrero-González, E. Romero-Cadaval, R.A. Lopes, J. Martins, Advantages of minimizing energy exchange instead of energy cost in prosumer microgrids, Energies 12 (2019), https://doi.org/10.3390/en12040719, https://www.mdpi.com/1996-1073/12/4/719.
[19] J. Widén, E. Wäckelgård, J. Paatero, P. Lund, Impacts of distributed photovoltaics on network voltages: Stochastic simulations of three Swedish low-voltage distribution grids, Electr. Power Syst. Res. 80 (2010) 1562–1571, https://doi.

org/10.1016/j.epsr.2010.07.007, https://www.sciencedirect.com/science/article/pii/S0378779610001707.
[20] F. Minelli, I. Ciriello, F. Minichiello, D. D'Agostino, From net zero energy buildings to an energy sharing model - the role of NZEBs in renewable energy communities, Renew. Energy (2024) 120110, https://doi.org/10.1016/j.renene.2024.120110, https://www.sciencedirect.com/science/article/pii/S0960148124001757.
[21] D. Coraci, S. Brandi, A. Capozzoli, Effective pre-training of a deep reinforcement learning agent by means of long short-term memory models for thermal energy management in buildings, Energy Convers. Manag. 291 (2023) 117303, https://doi.org/10.1016/j.enconman.2023.117303, https://www.sciencedirect.com/science/article/pii/S0196890423006490.
[22] S. Brandi, A. Gallo, A. Capozzoli, A predictive and adaptive control strategy to optimize the management of integrated energy systems in buildings, Energy Rep. 8 (2022) 1550–1567, https://doi.org/10.1016/j.egyr.2021.12.058, https://www.sciencedirect.com/science/article/pii/S2352484721014979.
[23] J.R. Vázquez-Canteli, S. Ulyanin, J. Kämpf, Z. Nagy, Fusing tensorflow with building energy simulation for intelligent energy management in smart cities, Sustain. Cities Soc. 45 (2019) 243–257, https://doi.org/10.1016/j.scs.2018.11.021, http://www.sciencedirect.com/science/article/pii/S2210670718314380.
[24] Y. Du, H. Zandi, O. Kotevska, K. Kurte, J. Munk, K. Amasyali, E. Mckee, F. Li, Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning, Appl. Energy 281 (2021) 116117, https://doi.org/10.1016/j.apenergy.2020.116117, http://www.sciencedirect.com/science/article/pii/S030626192031535X.
[25] K. Nweye, B. Liu, P. Stone, Z. Nagy, Real-world challenges for multi-agent reinforcement learning in grid-interactive buildings, Energy AI 10 (2022) 100202, https://doi.org/10.1016/j.egyai.2022.100202, https://www.sciencedirect.com/science/article/pii/S2666546822000489.
[26] K. Kaspar, M. Ouf, U. Eicker, A critical review of control schemes for demand-side energy management of building clusters, Energy Build. 257 (2022) 111731, https://doi.org/10.1016/j.enbuild.2021.111731, https://www.sciencedirect.com/science/article/pii/S037877882101015X.
[27] S. D'Oca, S.P. Corgnati, T. Buso, Smart meters and energy savings in Italy: Determining the effectiveness of persuasive communication in dwellings, Energy Res. Soc. Sci. 3 (2014) 131–142, https://doi.org/10.1016/j.erss.2014.07.015, https://www.sciencedirect.com/science/article/pii/S2214629614000930.
[28] E. Dal Cin, G. Carraro, G. Volpato, A. Lazzaretto, P. Danieli, A multi-criteria approach to optimize the design-operation of energy communities considering economic-environmental objectives and demand side management, Energy Convers. Manag. 263 (2022) 115677, https://doi.org/10.1016/j.enconman.2022.115677, https://www.sciencedirect.com/science/article/pii/S0196890422004733.
[29] A. Cosic, M. Stadler, M. Mansoor, M. Zellinger, Mixed-integer linear programming based optimization strategies for renewable energy communities, Energy 237 (2021) 121559, https://doi.org/10.1016/j.energy.2021.121559, https://www.sciencedirect.com/science/article/pii/S0360544221018077.
[30] J. Joe, P. Karava, A model predictive control strategy to optimize the performance of radiant floor heating and cooling systems in office buildings, Appl. Energy 245 (2019) 65–77, https://doi.org/10.1016/j.apenergy.2019.03.209, https://www.sciencedirect.com/science/article/pii/S0306261919306191.
[31] S. Brandi, M. Fiorentini, A. Capozzoli, Comparison of online and offline deep reinforcement learning with model predictive control for thermal energy management, Autom. Constr. 135 (2022) 104128, https://doi.org/10.1016/j.autcon.2022.104128, https://www.sciencedirect.com/science/article/pii/S0926580522000012.
[32] A. Hernandez-Matheus, M. Löschenbrand, K. Berg, I. Fuchs, M. Aragüés-Peñalba, E. Bullich-Massagué, A. Sumper, A systematic review of machine learning techniques related to local energy communities, Renew. Sustain. Energy Rev. 170 (2022) 112651, https://doi.org/10.1016/j.rser.2022.112651, https://www.sciencedirect.com/science/article/pii/S1364032122005433.
[33] W. Cai, A.B. Kordabad, S. Gros, Energy management in residential microgrid using model predictive control-based reinforcement learning and Shapley value, Eng. Appl. Artif. Intell. 119 (2023) 105793, https://doi.org/10.1016/j.engappai.2022.105793, https://www.sciencedirect.com/science/article/pii/S0952197622007837.
[34] F. Conte, F. D'Antoni, G. Natrella, M. Merone, A new hybrid AI optimal management method for renewable energy communities, Energy AI 10 (2022) 100197, https://doi.org/10.1016/j.egyai.2022.100197, https://www.sciencedirect.com/science/article/pii/S266654682200043X.
[35] A. Petrucci, G. Barone, A. Buonomano, A. Athienitis, Modelling of a multi-stage energy management control routine for energy demand forecasting, flexibility, and optimization of smart communities using a recurrent neural network, Energy Convers. Manag. 268 (2022) 115995, https://doi.org/10.1016/j.enconman.2022.115995, https://www.sciencedirect.com/science/article/pii/S0196890422007889.
[36] F. Zhou, Y. Li, W. Wang, C. Pan, Integrated energy management of a smart community with electric vehicle charging using scenario based stochastic model predictive control, Energy Build. 260 (2022) 111916, https://doi.org/10.1016/j.enbuild.2022.111916, https://www.sciencedirect.com/science/article/pii/S0378778822000871.
[37] G. Barone, G. Brusco, D. Menniti, A. Pinnarelli, G. Polizzi, N. Sorrentino, P. Vizza, A. Burgio, How smart metering and smart charging may help a local energy community in collective self-consumption in presence of electric vehicles, Energies 13 (2020), https://doi.org/10.3390/en13164163, https://www.mdpi.com/1996-1073/13/16/4163.
[38] P. Olivella-Rosell, F. Rullan, P. Lloret-Gallego, E. Prieto-Araujo, R. Ferrer-San-José, S. Barja-Martinez, S. Bjarghov, V. Lakshmanan, A. Hentunen, J. Forsström, S.Ø. Ottesen, R. Villafafila-Robles, A. Sumper, Centralised and distributed optimization for aggregated flexibility services provision, IEEE Trans. Smart Grid 11 (2020) 3257–3269, https://doi.org/10.1109/TSG.2019.2962269.
[39] L. Canese, G.C. Cardarilli, L. Di Nunzio, R. Fazzolari, D. Giardino, M. Re, S. Spanò, Multi-agent reinforcement learning: A review of challenges and applications, Appl. Sci. 11 (2021), https://doi.org/10.3390/app11114948, https://www.mdpi.com/2076-3417/11/11/4948.
[40] Z. Nagy, G. Henze, S. Dey, J. Arroyo, L. Helsen, X. Zhang, B. Chen, K. Amasyali, K. Kurte, A. Zamzam, H. Zandi, J. Drgoňa, M. Quintana, S. McCullogh, J.Y. Park, H. Li, T. Hong, S. Brandi, G. Pinto, A. Capozzoli, D. Vrabie, M. Bergés, K. Nweye, T. Marzullo, A. Bernstein, Ten questions concerning reinforcement learning for building energy management, Build. Environ. 241 (2023) 110435, https://doi.org/10.1016/j.buildenv.2023.110435, https://www.sciencedirect.com/science/article/pii/S0360132323004626.
[41] F. Charbonnier, T. Morstyn, M.D. McCulloch, Scalable multi-agent reinforcement learning for distributed control of residential energy flexibility, Appl. Energy 314 (2022) 118825, https://doi.org/10.1016/j.apenergy.2022.118825, https://www.sciencedirect.com/science/article/pii/S0306261922002689.
[42] Y. Qin, J. Ke, B. Wang, G.F. Filaretov, Energy optimization for regional buildings based on distributed reinforcement learning, Sustain. Cities Soc. 78 (2022) 103625, https://doi.org/10.1016/j.scs.2021.103625, https://www.sciencedirect.com/science/article/pii/S2210670721008891.
[43] R. Lowe, Y.I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, I. Mordatch, Multi-agent actor-critic for mixed cooperative-competitive environments, Adv. Neural Inf. Process. Syst. 30 (2017).
[44] I. Jendoubi, F. Bouffard, Multi-agent hierarchical reinforcement learning for energy management, Appl. Energy 332 (2023) 120500, https://doi.org/10.1016/j.apenergy.2022.120500, https://www.sciencedirect.com/science/article/pii/S0306261922017573.
[45] A. Pigott, C. Crozier, K. Baker, Z. Nagy, Gridlearn: Multiagent reinforcement learning for grid-aware building energy management, Electr. Power Syst. Res. 213 (2022) 108521, https://doi.org/10.1016/j.epsr.2022.108521, https://www.sciencedirect.com/science/article/pii/S0378779622006320.
[46] K. Nweye, S. Sankaranarayanan, Z. Nagy, Merlin: Multi-agent offline and transfer learning for occupant-centric operation of grid-interactive communities, Appl. Energy 346 (2023) 121323, https://doi.org/10.1016/j.apenergy.2023.121323, https://www.sciencedirect.com/science/article/pii/S0306261923006876.
[47] K. Nweye, K. Kaspar, G. Buscemi, G. Pinto, H. Li, T. Hong, M. Ouf, A. Capozzoli, Z. Nagy, A framework for the design of representative neighborhoods for energy flexibility assessment in Citylearn, J. Build. Perform. Simul. (2023), Building Simulation 23.
[48] A. Gallo, M.S. Piscitelli, L. Fenili, A. Capozzoli, RECsim—virtual testbed for control strategies implementation in renewable energy communities, in: J. Littlewood, R.J. Howlett, L.C. Jain (Eds.), Sustainability in Energy and Buildings 2022, Springer Nature Singapore, Singapore, 2023, pp. 313–323.
[49] Z. Wang, B. Chen, H. Li, T. Hong, AlphaBuilding ResCommunity: A multi-agent virtual testbed for community-level load coordination, Adv. Appl. Energy 4 (2021) 100061, https://doi.org/10.1016/j.adapen.2021.100061, https://www.sciencedirect.com/science/article/pii/S2666792421000536.
[50] I. Staffell, D. Brett, N. Brandon, A. Hawkes, A review of domestic heat pumps, Energy Environ. Sci. 5 (2012) 9291–9306, https://doi.org/10.1039/C2EE22653G.
[51] I. Richardson, M. Thomson, D. Infield, C. Clifford, Domestic electricity use: A high-resolution energy demand model, Energy Build. 42 (2010) 1878–1887, https://doi.org/10.1016/j.enbuild.2010.05.023, https://www.sciencedirect.com/science/article/pii/S0378778810001854.
[52] E. McKenna, M. Thomson, High-resolution stochastic integrated thermal–electrical domestic demand model, Appl. Energy 165 (2016) 445–461, https://doi.org/10.1016/j.apenergy.2015.12.089, https://www.sciencedirect.com/science/article/pii/S0306261915016621.
[53] demod, https://demod.readthedocs.io/en/latest/index.html, 2022.
[54] pvlib python, https://pvlib-python.readthedocs.io/en/stable/, 2022.
[55] J.R. Vazquez-Canteli, S. Dey, G. Henze, Z. Nagy, Citylearn: Standardizing research in multi-agent reinforcement learning for demand response and urban energy management, https://arxiv.org/abs/2012.10504, arXiv:2012.10504, 2020.
[56] A. Amato, M. Bilardo, E. Fabrizio, V. Serra, F. Spertino, Energy evaluation of a PV-based test facility for assessing future self-sufficient buildings, Energies 14 (2021), https://doi.org/10.3390/en14020329, https://www.mdpi.com/1996-1073/14/2/329.
[57] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, second ed., The MIT Press, 2018, http://incompleteideas.net/book/the-book-2nd.html.
[58] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, Playing Atari with deep reinforcement learning, arXiv:1312.5602, 2013.
[59] A. Hill, A. Raffin, M. Ernestus, A. Gleave, A. Kanervisto, R. Traore, P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, Y. Wu, Stable baselines, https://github.com/hill-a/stable-baselines, 2018.
[60] A. Standard, Thermal environmental conditions for human occupancy, ANSI/ASHRAE 55 (5) (2017).
[61] Arera, https://www.arera.it/, 2022.